A GUIDE TO ECONOMETRICS

SIXTH EDITION

PETER KENNEDY
Simon Fraser University

Blackwell Publishing

frequently merely a euphemism for obtaining plausible numbers to provide ceremonial adequacy for a theory chosen and defended on *a priori* grounds." For a completely opposite cynical view, Blaug (1980, p. 257) feels that econometricians "express a hypothesis in terms of an equation, estimate a variety of forms for that equation, select the best fit, discard the rest, and then adjust the theoretical argument to rationalize the hypothesis that is being tested."

- It should be borne in mind that despite the power, or lack thereof, of hypothesis tests, often conclusions are convincing to a researcher only if supported by personal experience. Nelson (1995, p. 141) captures this subjective element of empirical research by noting that "what often really seems to matter in convincing a male colleague of the existence of sex discrimination is not studies with 10 000 'objective' observations, but rather a particular single direct observation: the experience of his own daughter."
- Hypothesis tests are usually conducted using a type I error rate (probability of rejecting a true null) of 5%, but there is no good reason why 5% should be preferred to some other percentage. The father of statistics, R. A. Fisher, suggested it in an obscure 1923 paper, and it has been blindly followed ever since. Rosnow and Rosenthal (1989, p. 1277) recognize that "surely, God loves the .06 as much as the .05." By increasing the type I error rate, the type II error rate (the probability of accepting the null when it is false) is lowered, so the choice of type I error rate should be determined by the relative costs of the two types of error, but this issue is usually ignored by all but Bayesians (see chapter 14). The .05 is chosen so often that it has become a tradition, prompting Kempthorne and Doerfler (1969, p. 231) to opine that "statisticians are people whose aim in life is to be wrong 5% of the time!"
- Most hypothesis tests fall into one of two categories. Suppose we are testing the null that the slope of x in a regression is zero. One reason we are doing this could be that we are genuinely interested in whether this slope is zero, perhaps because it has some substantive policy implication or is crucial to supporting some economic theory. This is the category for which the traditional choice of a 5% type I error rate is thought to be applicable. But it may be that we have no real interest in this parameter and that some other parameter in this regression is of primary interest. In this case, the reason we are testing this null hypothesis is because if we fail to reject it we can drop this explanatory variable from the estimating equation, thereby improving estimation of this other parameter. In this context, the traditional choice of 5% for the type I error is no longer an obvious choice, something that is not well recognized by practitioners. As explained in chapter 6, omitting a relevant explanatory variable in general causes bias. Because most econometricians fear bias, they need to be very careful that they do not drop an explanatory variable that belongs in the regression. Because of this they want the power of their test (the probability of rejecting the null when it is false) to be high, to ensure that they do not drop a relevant explanatory variable. But choosing a low type I error, such as 5%, means that power will be much lower than if a type I error of, say, 30% was chosen. Somehow the type I error needs to be chosen so as to maximize the quality of the estimate of the parameter of primary interest. Maddala and Kim (1998, p. 140) suggest a type I error of about 25%. Further discussion of this important practical issue occurs in the general notes to section 5.2, in section 6.2 and its general notes, and in the technical notes to section 13.5.
- For a number of reasons, tests of significance can sometimes be misleading. A good discussion can be found in Bakan (1966). One of the more interesting problems in this respect is the fact that almost any parameter can be found to be significantly different from zero if the sample size is sufficiently large. (Almost every relevant independent variable will have *some* influence, however small, on a dependent variable; increasing the sample size will reduce the variance and eventually make this influence "statistically significant.") Thus, although a researcher wants a large sample size to generate more accurate estimates, too large a sample size might cause difficulties in interpreting the usual tests of significance.

the traditional
s ght to be
e ...ve no real
iat some other
primary inter-
are testing this
fail to reject it
riable from the
roving estima-
his context, the
ype I error is no
hing that is not
As explained in
xplanatory vari-
use most econo-
be very careful
ory variable that
use of this they
ie probability of
e) to be high, to
vant explanatory
e I error, such as
uch lower than if
hosen. Somehow
en so as to maxi-
of the parameter
ar Kim (1998,
r about 25%.
portant practical
to section 5.2, in
, and in the tech-

s of significance
good discussion
One of the more
ect is the fact that
und to be signifi-
sample size is suf-
levant independent
ce, however small,
easing the sample
e and eventually
ally significant.")
vants a large sam-
rate estimates, too
cause difficulties
s of significance.

McCloskey and Ziliak (1996) look carefully at a large number of empirical studies in economics and conclude that researchers seem not to appreciate that statistical significance does not imply economic significance. One must ask if the magnitude of the coefficient in question is large enough for its explanatory variable to have a meaningful (as opposed to "significant") influence on the dependent variable. This is called the *too-large sample size problem*. One suggestion for dealing with this problem is to report *beta coefficient* estimates – scale the usual coefficient estimates so that they measure the number of standard deviation changes in the dependent variable due to a standard deviation change in the explanatory variable. A second suggestion is to adjust the significance level downward as the sample size grows; for a formalization see Leamer (1978, pp. 88–9, 104–5). See also Attfield (1982). Leamer would also argue (1988, p. 331) that this problem would be resolved if researchers recognized that genuinely interesting hypotheses are neighborhoods, not points. Another interesting dimension of this problem is the question of what significance level should be employed when replicating a study with new data; conclusions must be drawn by considering both sets of data as a unit, not just the new set of data. For discussion see Busche and Kennedy (1984). Another interesting example in this context is the propensity for published studies to contain a disproportionately large number of type I errors; studies with statistically significant results tend to get published, whereas those with insignificant results do not. For comment see Feige (1975). Yet another example that should be mentioned here is pretest bias, discussed in chapter 13.

- In psychometrics these problems with significance testing have given rise to a book entitled "What if there were no significance tests?" (Harlow, Mulaik, and Steiger, 1997) and journal policies not to publish papers that do not report effect size (the magnitude of a treatment's impact, usually measured in terms of standard deviations of the phenomenon in question). Loftus's (1993, p. 250) opinion that "hypothesis testing is overstated, overused and practically useless as a means of illuminating what the data in some experiment are trying to tell us," is shared by many. Nester (1996) has a collection of similar quotes berating significance testing. One way of alleviating this problem is to report confidence intervals rather than hypothesis test results; this allows a reader to see directly the magnitude of the parameter estimate along with its uncertainty.

  In econometrics, McCloskey (1998, chapter 8) summarizes her several papers on the subject, chastising the profession for its tendency to pay undue homage to significance testing. McCloskey and Ziliak (1996, p. 112) cogently sum up this view as follows:

  > No economist has achieved scientific success as a result of a statistically significant coefficient. Massed observations, clever common sense, elegant theorems, new policies, sagacious economic reasoning, historical perspective, relevant accounting: these have all led to scientific success. Statistical significance has not.

  Ziliak and McCloskey (2004) is an update of their earlier study, finding that researchers continue to abuse significance tests; this paper is followed by a set of interesting commentaries.

- Tukey (1969) views significance testing as "sanctification" of a theory, with a resulting unfortunate tendency for researchers to stop looking for further insights. Sanctification via significance testing should be replaced by searches for additional evidence, both corroborating evidence, and, especially, disconfirming evidence. If your theory is correct, are there testable implications? Can you explain a range of interconnected findings? Can you find a bundle of evidence consistent with your hypothesis but inconsistent with alternative hypotheses? Abelson (1995, p. 186) offers some examples. A related concept is encompassing: Can your theory encompass its rivals in the sense that it can explain other models' results? See Hendry (1988).
- Inferences from a model may be sensitive to the model specification, the validity of which may be in doubt. A *fragility analysis* is recommended to deal with this; it examines the range of inferences resulting from the range of believable model

Chapter 22

Applied Econometrics

22.1 Introduction

The preceding chapters have focused on econometric theory. Unfortunately, unpleasant realities of real-world data force applied econometricians to violate the prescriptions of econometric theory as taught by our textbooks. Leamer (1978, p. vi) vividly describes this behavior as wanton sinning:

> As it happens, the econometric modeling was done in the basement of the building and the econometric theory courses were taught on the top floor (the third). I was perplexed by the fact that the same language was used in both places. Even more amazing was the transmogrification of particular individuals who wantonly sinned in the basement and metamorphosed into the highest of high priests as they ascended to the third floor.

It is no secret that there is a world of difference between applied and theoretical econometrics. In fact, there is a remarkable lack of communication between econometric theorists and applied econometricians – the former, who are often called upon to teach applied econometrics courses, are notorious for teaching econometric theory in these courses. (Examples are given, and an applied paper is usually required, to justify calling the course an applied econometrics course!)

In these "applied" courses students typically are taught, in hands-on fashion, how to undertake a wide variety of econometric techniques. Examples at the elementary level are the use and interpretation of dummy variables, the logic of F and chi-square tests, and testing and correcting for nonspherical errors. Examples at a more advanced level are testing for unit roots and cointegration, correcting for sample selection bias, performing Hausman tests, and estimating using Tobit, Poisson, and ordered probit models. But the focus is on the mechanics of estimation and testing rather than on the fundamentals of applied work such as problem articulation, data cleaning, and model specification. In short, teaching is technique oriented rather than problem oriented.